Role of stem-cell divisions in cancer risk

ARISING FROM S. Wu, S. Powers, W. Zhu & Y. A. Hannun *Nature* **529,** 43–47 (2016); doi:10.1038/nature16166

We recently reported a strong correlation between the incidence of cancers and the number of stem-cell divisions in the corresponding normal tissues[1]. We interpreted this correlation to mean that random genomic alterations (termed replicative or intrinsic) arising during DNA replication, as well as mutations that arise owing to environmental (extrinsic) and inherited factors, have important roles in tumorigenesis[1–3]; however, we did not quantify the contribution of intrinsic versus extrinsic factors to any cancer type[1]. In their study, Wu *et al.*[4] estimated an upper bound for the contribution of intrinsic factors to many cancer types, concluding that intrinsic factors account for "less than 10–30%" of cancer cases. We believe that several of the assumptions made by these authors led them to underestimate the role of intrinsic factors, and we further show that one of their methods leads to the conclusion that extrinsic factors account for >85% of the risk in situations in which extrinsic factors have no role. There is a Reply to this Comment by Wu, S., Zhu, W. & Hannun, Y. A. *Nature* **548,** http://dx.doi.org/10.1038/nature23303 (2017).

On the basis of the data in ref. 1, Wu *et al.* chose seven cancer types to obtain a 'lower bound intrinsic risk' (LBIR) line, representing cancers that they assumed to carry a zero extrinsic risk (figure 3a in ref. 4). Instead of using the cancer types at the bottom of figure 3a in ref. 4 to define an LBIR line, one could analogously use cancer types at the top of figure 3a in ref. 4 to define an 'upper bound extrinsic risk' (UBER) line (Methods). 74% of the seven cancer types nearest this line are preventable (http://www.cancerresearchuk.org/cancer-info/cancerstats/causes/preventable) and 90% would therefore be a conservative estimate for the extrinsic risk in the cancers on this line. Using this UBER line to determine the risk of the 24 other cancer types, on average, 80% of the total risk can be calculated to result from intrinsic factors (Fig. 1). In contrast, using the LBIR line, on average, only 7% of the total risk is calculated to result from intrinsic risks (extended data table 1 in ref. 4). If one assumes values of extrinsic risk for the UBER line that are inconsistent with epidemiologic evidence (for example, extrinsic risk <25% or ≥99%), the risks attributable to intrinsic factors are either extremely high or extremely low (Fig. 1). Boundary-based approaches can therefore yield widely variant conclusions simply depending on whether an upper or lower boundary is chosen and the fraction of extrinsic risk the boundary is assumed to represent.

The estimates for the lifetime number of stem-cell divisions are noisy given the many different and complex biologic experiments required for their determination. In our original paper, we performed a robustness analysis by assessing the effect of noise on all cancer types analysed[1]. Because the lifetime number of stem-cell divisions in each organ is critical for the definition of the LBIR line, and for the distance of a cancer type from the LBIR line, noise in the stem cell estimates could strongly affect estimates of extrinsic risk. Wu *et al.* recognized this problem and performed a robustness analysis. However, we believe that the effects of noise were not taken sufficiently into account for defining the LBIR line; this line forms the baseline to which all other cancer types are compared. We performed simulations to evaluate the effect of noise on the data used to obtain the LBIR line and on the conclusions reached by Wu *et al.* (Methods). A typical simulation in which the extrinsic risk was assumed to be 10% is shown in Fig. 2a. The introduced noise transforms the green dots, representing the true values, into the red dots. The positions of all the red dots not on the green line in Fig. 2a are artefacts of noise. The histogram in Fig. 2b shows the aggregate results of 10,000 such simulations. The LBIR approach overestimated the fraction of risk attributable to extrinsic factors; the median extrinsic risk was estimated to be 86% in these 10,000 simulations, although the assumed true risk was 10%. In fact, the median extrinsic risk was incorrectly estimated at 86% of the total risk, regardless of whether the true extrinsic risk was 0%, 100%, or any value in between.

Our reading of the methods adopted by Wu *et al.* indicates that they assumed a linear relationship between cancer incidence and stem-cell divisions among cancer types with the same extrinsic risk. A corollary of this assumption is that cancer incidence in a tissue can be determined exclusively by the number of stem-cell divisions in the tissue and the degree of extrinsic risk. We do not believe that this assumption of linearity is justified because it disregards other factors that could influence incidence, such as the number of required mutations. Suppose, for example, that cancer types 'A' and 'B' have identical extrinsic risks and identical stem-cell divisions. The linearity assumption mandates that the incidences of cancer types A and B would be identical. However, figure 4 of ref. 4 demonstrates that if cancer type A requires two mutations to progress to malignancy, and cancer type B requires three mutations, then the incidence of cancer type A will be orders of magnitude higher than cancer type B.

We believe that three other approaches used by Wu *et al.* to find a bound for the role of intrinsic risk were non-conservative, as explained in the following. First, in their analysis of Surveillance, Epidemiology, and End Results Program (SEER) cancer incidence data, Wu *et al.* assumed that all of the variation above the lowest value was due to extrinsic factors. But it has been documented that these incidences are affected by a variety of biological and methodological factors that are not linked to extrinsic risk—as well as by omnipresent noise (refs 5–9 and http://www.cancer.gov/research/progress/snapshots/kidney).

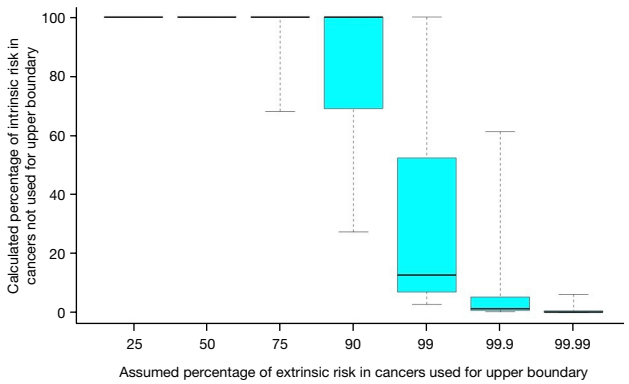

Figure 1 | The effect of different assumptions of extrinsic risk for the upper boundary line. For the same 31 cancer types in figure 3a of ref. 4, an 'upper boundary line' was defined by cancer types with the 'highest' relative risk among the 31 cancer types (Methods). Various estimates of extrinsic risk (*x* axis) on the upper boundary line were then used to calculate the fraction of intrinsic risk of the other 24 cancer types. The box plots represent the median (black bar within the box), interquartiles (blue box), and extreme values (black bars outside the box) of calculated intrinsic risks of these other cancer types (*y* axis). The calculated fractions of intrinsic risks are very sensitive to the degree of risk that the boundary is assumed to represent.

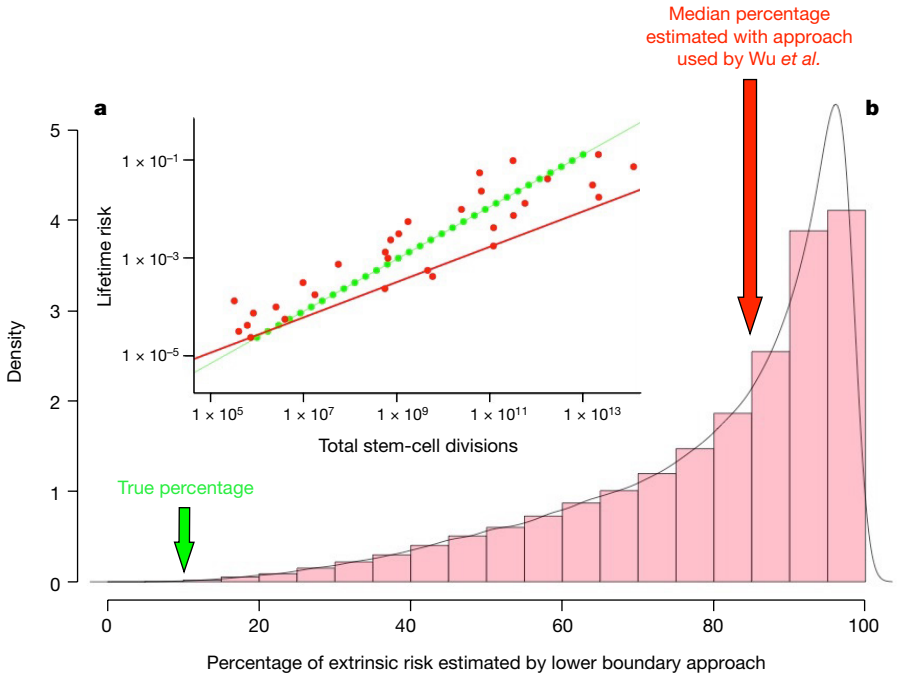

Figure 2 | The effect of noise in stem cell estimates of cancer risk. a, 31 hypothetical cancer types are considered, with the total number of stem-cell divisions ranging across the values used in refs 1, 4. For each cancer type, 10% of the risk results from extrinsic factors and the remainder from intrinsic factors. The lifetime risks for each cancer type, representing the sum of extrinsic and intrinsic risks, are represented by the green dots. When noise is introduced into the stem cell estimates for each of the 31 cancer types, the red dots are obtained (Methods). The positions of all the red dots, including those near the red line, are artefacts of noise; the true values are those of the green dots. **b**, Ten thousand simulations exactly like the one shown in **a** were performed (Methods). The proportions of extrinsic risk were calculated as in Wu *et al.*[4] and graphed as a histogram (pink bars) or as a density (black line). The true value for all cancer types is 10% (green arrow), whereas the LBIR approach used by Wu *et al.* estimates a median extrinsic risk of 86% (red arrow).

Second, in their analysis of signatures, it appeared as though Wu *et al.* assumed that signatures not obviously associated with ageing were exclusively due to extrinsic factors. However, the investigators who discovered these signatures stated in their study that "The mechanistic basis of some signatures is, at least partially, understood but for many it remains speculative or unknown"[10]. Finally, in their mathematical modelling approach, Wu *et al.* assumed that clonal expansions have no effect on the acquisition of driver genes and that all cancer types have the same number of driver genes. In our view, these three approaches, when based on non-conservative assumptions such as used by Wu *et al.*, cannot be used to determine reliable bounds.

Methods

Random noise was incorporated into the estimates of lifetime stem-cell divisions by assuming uniform distributions centred on the literature estimates plus or minus two orders of magnitude, as in ref. 1. Quantile regressions centred at the 12.5 percentile were performed in each of 10,000 simulations to derive the intrinsic risk (lower boundary) lines. The proportion of extrinsic risk for each cancer type not in the lowest quartile was then computed, as in ref. 4, and the overall distribution of proportions shown in Fig. 2b. Quantile regressions centred at the 87.5 percentile were performed for deriving the upper boundary lines (Fig. 1).

Data availability. All data are available from the corresponding author upon reasonable request.

Cristian Tomasetti[1,2], Rick Durrett[3], Marek Kimmel[4], Amaury Lambert[5], Giovanni Parmigiani[6], Ann Zauber[7] & Bert Vogelstein[8]

[1]Division of Biostatistics and Bioinformatics, Department of Oncology, Sidney Kimmel Cancer Center, Johns Hopkins University School of Medicine, 550 North Broadway, Baltimore, Maryland 21205, USA.
email: ctomasetti@jhu.edu
[2]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland 21205, USA.
[3]Department of Mathematics, Duke University, Durham, North Carolina 27708, USA.
[4]Department of Statistics, Rice University, Houston, Texas 77005, USA.
[5]Laboratoire de Probabilités & Modéles Aléatoires, UMPC Université Paris 06, 75252 Paris, France.
[6]Department of Biostatistics, Harvard T.H. Chan School of Public Health & Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA.
[7]Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York 10017, USA.
[8]Ludwig Center & Howard Hughes Medical Institute, Johns Hopkins Kimmel Cancer Center, 1650 Orleans Street St, Baltimore, Maryland 21205, USA.

Received 20 January 2016; accepted 13 April 2017.

1. Tomasetti, C. & Vogelstein, B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* **347,** 78–81 (2015).
2. Tomasetti, C. & Vogelstein, B. Musings on the theory that variation in cancer risk among tissues is explained by the number of divisions of the normal stem cells. Preprint at https://arxiv.org/abs/1501.05035 (2015).
3. Tomasetti, C. & Vogelstein, B. Cancer risk: role of environment—response. *Science* **347,** 729–731 (2015).
4. Wu, S., Powers, S., Zhu, W. & Hannun, Y. A. Substantial contribution of extrinsic risk factors to cancer development. *Nature* **529,** 43–47 (2016).
5. Lansdorp-Vogelaar, I. *et al.* Contribution of screening and survival differences to racial disparities in colorectal cancer rates. *Cancer Epidemiol. Biomarkers Prev.* **21,** 728–736 (2012).
6. Davies, L. & Welch, H. G. Increasing incidence of thyroid cancer in the United States, 1973-2002. *J. Am. Med. Assoc.* **295,** 2164–2167 (2006).
7. Edwards, B. K. *et al.* Annual report to the nation on the status of cancer, 1975-2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. *Cancer* **116,** 544–573 (2010).
8. Greiner, T. C., Medeiros, L. J. & Jaffe, E. S. Non-Hodgkin's lymphoma. *Cancer* **75** (Suppl.), 370–380 (1995).
9. Kort, E. J., Paneth, N. & Vande Woude, G. F. The decline in U.S. cancer mortality in people born since 1925. *Cancer Res.* **69,** 6500–6505 (2009).
10. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500,** 415–421 (2013).

Author Contributions C.T. and B.V. conceived and designed the research. C.T., R.D., M.K., A.L., G.P., A.Z. and B.V. performed related research. C.T. and B.V. wrote the first draft of the article and all authors contributed to the submitted version.

Competing Financial Interests Declared none.

doi:10.1038/nature23302

Wu *et al.* reply

REPLYING TO C. Tomasetti *et al. Nature* **548,** http://dx.doi.org/10.1038/nature23302 (2017)

In the accompanying Comment[1], Tomasetti *et al.* consider the tumorigenic effects of both 'random mutations' (intrinsic) arising during DNA replication, as well as mutations that occur owing to environmental (extrinsic) and inherited factors, driving the discussion into the domain of estimating the contributions of extrinsic factors alongside intrinsic/unmodifiable factors. Originally, Tomasetti and Vogelstein estimated the contribution of intrinsic factors as 64% on the basis of the correlation between stem-cell division and lifetime cancer risk[2]. However, our thought experiment (figures 1 and 2 in our study[3]) showed that this correlation does not distinguish the effects of extrinsic versus intrinsic mutagens acting at the level of cell division. We further provided four distinct approaches to estimate the contribution of extrinsic factors, and they all converge on an estimate of 70–90% (that is, a contribution of intrinsic factors at 10–30%).

Tomasetti *et al.* state that the seven cancer types we used to define the 'intrinsic' risk line have 'zero intrinsic risk'. However, in our study, we stated that the 'intrinsic' risk lines themselves represent an upper estimate of intrinsic risk, allowing that these cancers are likely to have extrinsic components. Tomasetti *et al.* further state that in our robustness analyses, we did not consider noise of stem-cell divisions for these cancers. However, we added noise estimates to all cancer types, including those seven cancers. Indeed, we also observed that the same seven cancers were also the same as those defining the 'intrinsic' risk line in most simulation cases[3].

Tomasetti *et al.* argue that if the seven cancer types at the top of our figure 3a were used to define an upper boundary line and assume an extrinsic risk of 90% for the upper boundary line, an averaged 80% total risk could be attributed to intrinsic factors. However, as per our analyses, that assumption would lead to more than half of the cancer types, including those known to have substantial extrinsic risks, to show negative extrinsic risks. This implies that either the regression for the upper boundary line or the assumption of extrinsic risk of 90% for the upper boundary line is unfounded. If we adopt the approach of Tomasetti *et al.* and use the upper boundary line but associate 99.9% extrinsic risk (consisting of cancers known to be nearly exclusively induced by known extrinsic factors), the majority of cancer types still remain above the 90% extrinsic risk line, in agreement with our conclusions.

Tomasetti *et al.* performed simulations to evaluate the effect of noise and conclude that our lower-boundary approach overestimated the extrinsic risk. However, some of their simulation settings could lead to erroneous conclusions. In particular, in their simulation, they assume that the extrinsic risks of all cancers is 10%; however, this contradicts their simultaneous use of a regression slope of 0.52 between $\log_{10}$(cancer risk) and $\log_{10}$(stem-cell division), because they derived the slope of 0.52 from their previous data, in which many cancers are already known to have substantial extrinsic risks (more than 10%). Indeed, as shown in figure 3a of our study[3], the estimated slope for cancers with relatively more intrinsic risk is 0.27 (the intrinsic risk line), that is, considerably different to 0.52 on the $\log_{10}$ scale. Thus, we feel that owing to potentially erroneous assumptions contradictory to the observed data, their simulation cannot be used to dispute our method.

In their Comment[1], Tomasetti *et al.* mention that we assumed that there is a linear relationship between cancer incidence and stem-cell divisions among cancer types with the same extrinsic risk. However, we did not make that assumption, and the key assumption that we did make was direct and biologically based: cancers with the same number of stem-cell divisions should share the same intrinsic cancer risk if the relationship between total stem-cell division and cancer risk is causal. Therefore, for any two cancers with the same total stem-cell division, the one with the higher incidence of cancer must represent the contribution of extrinsic risk.

Tomasetti *et al.* raised further concerns regarding the other approaches we used. Although we agree with some of these points, such as incorporating clonal dynamics into future modelling for more accurate estimates, we cannot agree that ours are faulty because of overly liberal assumptions. The clonal expansion issue was partly addressed in our model that assumes every tissue cell to be a stem cell (figure 4b in our study[3]), which can be viewed as clonal expansion to the tissue size at the very early stage. Under this conservative assumption, the theoretical intrinsic risks are still found to be quite low. Estimation of extrinsic risks from mutational signatures is also conservative as extrinsic factors may cause cancers through many avenues. We realize that each approach has its own limitations, which led us to employ four independent approaches, each of which showed high concordance.

Author S. Powers was not available to work on this Reply.

Song Wu[1,2], Wei Zhu[1,2] & Yusuf A. Hannun[2,3,4,5]

[1]Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, New York 11794, USA.
[2]Stony Brook Cancer Center, Stony Brook University, Health Sciences Center, Stony Brook, New York 11794, USA.
email: Yusuf.Hannun@stonybrookmedicine.edu
[3]Department of Pathology, Stony Brook University, Health Sciences Center, Stony Brook, New York 11794, USA.
[4]Department of Medicine, Stony Brook University, Health Sciences Center, Stony Brook, New York 11794, USA.
[5]Department of Biochemistry, Stony Brook University, Health Sciences Center, Stony Brook, New York 11794, USA.

1. Tomasetti, C. *et al.* Role of stem-cell divisions in cancer risk. *Nature* **548,** http://dx.doi.org/10.1038/nature23302 (2017).
2. Tomasetti, C. & Vogelstein, B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* **347,** 78–81 (2015).
3. Wu, S., Powers, S., Zhu, W. & Hannun, Y. A. Substantial contribution of extrinsic risk factors to cancer development. *Nature* **529,** 43–47 (2016).

doi:10.1038/nature23303